Quality Summarization

recommendation on biometric quality summarization across the application domain

Elham Tabassi + Patrick Grother

July 2007

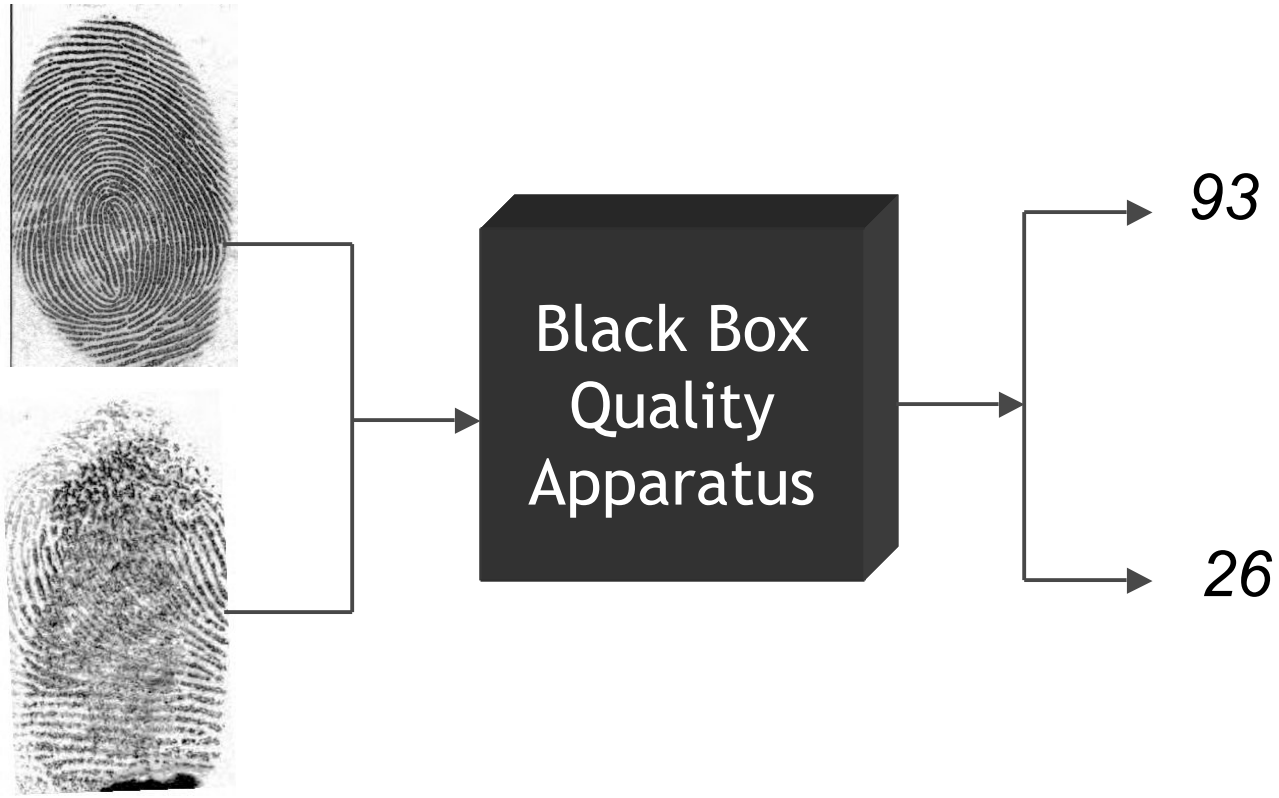
motivation
performance related quality monitoring

quality summarization supports monitoring

- over time (to expose seasonal variation, or trends),
- for each sensor (to identify defective devices),
- at each site (to identify problem locations)
- of officials or attendants (to assess adherence to operating procedures), and
- per user basis (to identify users that consistently yield low quality samples

In each case the quality summaries can be used to identify departures from the application specific historical norms, or design targets.

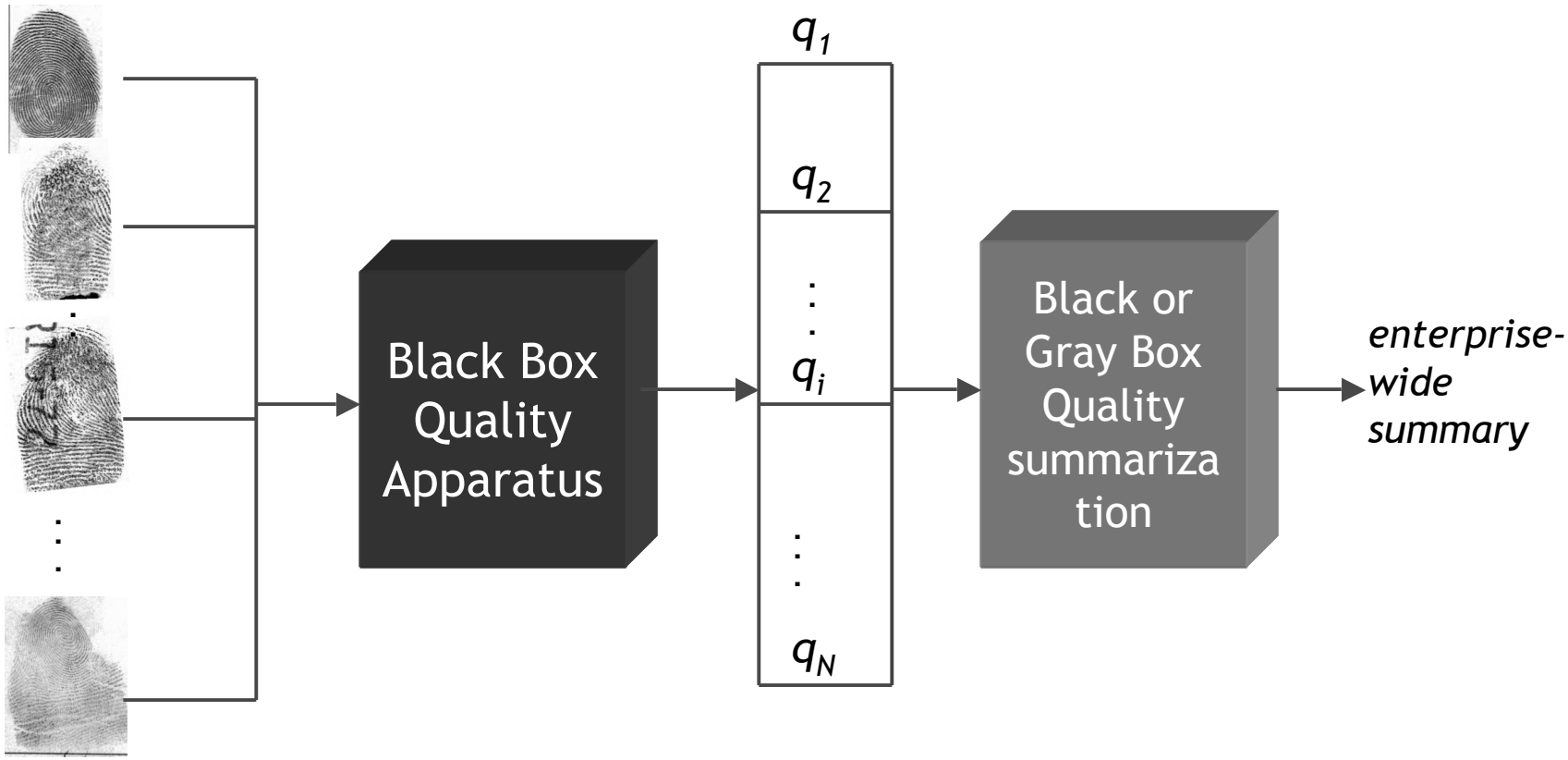
biometric sample quality



A biometric quality assessment method derives a numerical quality value from an input biometric sample. The quality value is related to the biometric error rates that are likely to be realized when the sample is matched.

enterprise-wide quality values

summarizing quality values computed across all retained samples in an enterprise into a single quality value representing the overall quality of the enterprise.



performance related quality summarization

- is not well represented by arithmetic mean of quality values
- should be a weighted average of the native quality values
  - weights for each quality level q (u_q) should be directly related to the error rate observed for samples of quality q
- should be on the range [0,100]
  - as specified in ISO/IEC 19784-1 BioAPI and ISO/IEC 29794-1
- could be the result of a biometric quality assessment calibration process
  - conducted by the provider, or by a third party laboratory
- should be performed across similar usage
  - e.g. quality summarization over all enrollment samples of an enterprise or per user basis in time and attendance applications

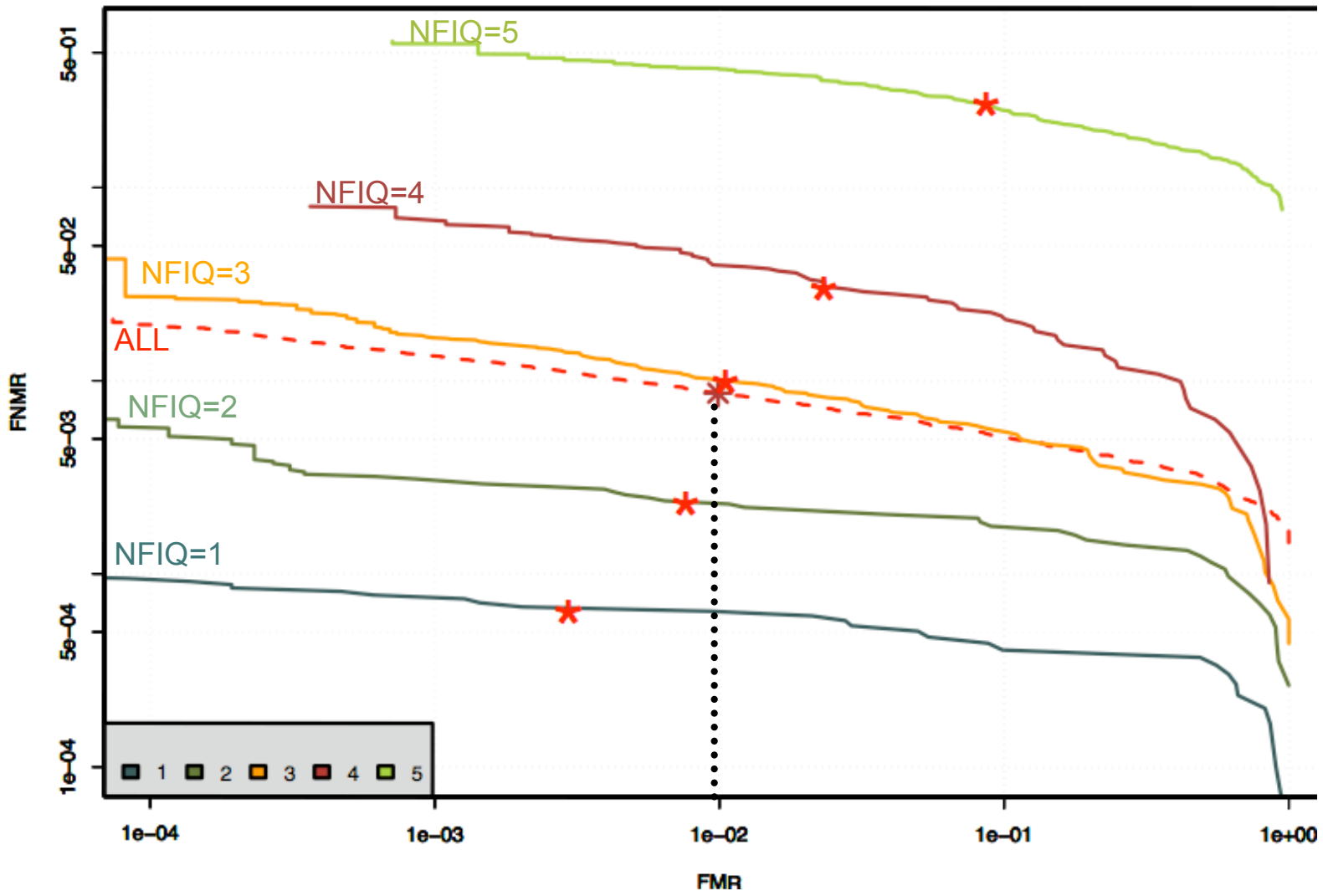
Best practice NFIQ summarization

$$Q = 102.75 - 2.75p_1 - 5.37p_2 - 14.38p_3 - 42.25p_4 - 102.75p_5$$

- p_i is the proportion of the fingerprints with NFIQ values i =1, 2, .., 5 in the enterprise.

- the weights reflect the likelihood that an observed false non-match involved a fingerprint of quality i
  - Weights were estimated using the observed false non-match rates from a set of leading commercial matching algorithms computed at some fixed threshold

NFIQ::(fmr,fnmr) at fixed threshold



computation of weights

1. for all V matching algorithms compute false non-match rate for each L quality levels at threshold τ

for $(v = 1, \ldots, V)$
  for $(i = 1, \ldots, L)$

$$\mathrm{FNMR}^v(\tau, i) = \frac{\left|\left\{s_{jj}^{(v)} : \; s_{jj} \le \tau, \;\; q_j^{(1)} \ge i, \; q_j^{(2)} = i\right\}\right|}{\left|\left\{s_{jj}^{(v)} : \; s_{jj} \le \infty, \; q_j^{(1)} \ge i, \; q_j^{(2)} = i\right\}\right|}$$

  end
end
which results in the following array

$$\begin{pmatrix} \mathrm{FNMR}^1(\tau,1) & \mathrm{FNMR}^2(\tau,1) & \ldots & \mathrm{FNMR}^V(\tau,1) \\ \mathrm{FNMR}^1(\tau,2) & \mathrm{FNMR}^2(\tau,2) & \ldots & \mathrm{FNMR}^V(\tau,2) \\ \ldots & \ldots & \ldots & \ldots \\ \mathrm{FNMR}^1(\tau,L) & \mathrm{FNMR}^2(\tau,L) & \ldots & \mathrm{FNMR}^V(\tau,L) \end{pmatrix}$$

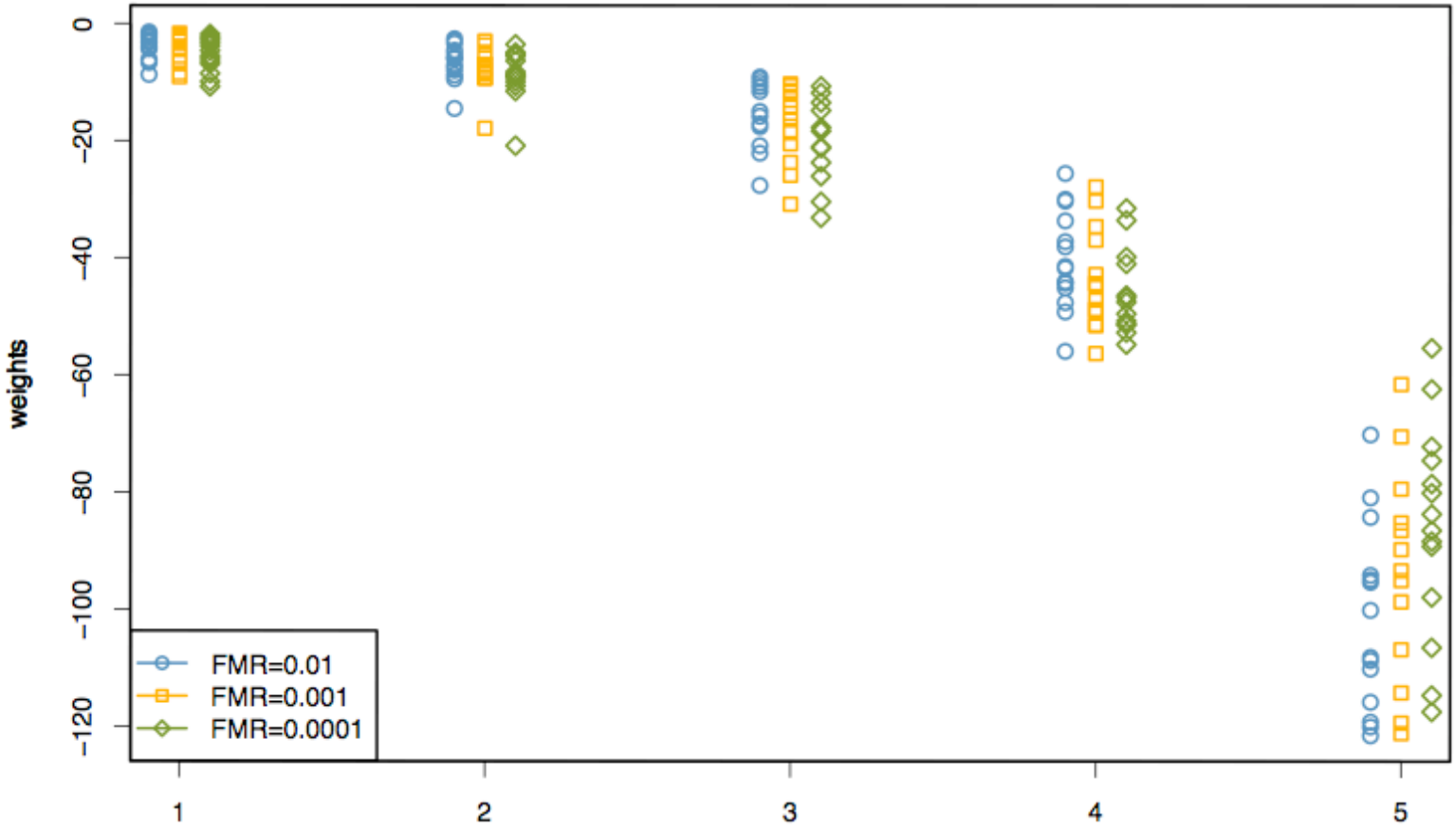
2. u_i *i = 1, …, L* is normalized false non-match

$$u_i = \frac{\sum_{v=1}^{V} \mathrm{FNMR}^v(\tau, i)}{\sum_{q=1}^{L} \sum_{v=1}^{V} \mathrm{FNMR}^v(\tau, q)}$$

3. Map (e.g. linear mapping of $[u_1, \ldots u_L]$ to [0,…100], for example NFIQ summarization is given by:

$$\bar{Q} = \frac{100 u_5}{u_5 - u_1} - \sum_{i=1}^{5} \frac{100 u_i}{u_5 - u_1} p_i$$

NFIQ summarization::dependence on operation threshold



Weights for NFIQ values 1, and 2 are quite robust to variation of the computing thresholds. Thresholds are set at overall false-match rates of 0.01, 0.001, and 0.0001. Each point corresponds to NFIQ weight estimated using similarity scores of a commercial matching algorithm on large operational fingerprint datasets.

dedicated weights for NFIQ summarization

Dedicated weights should be computed in verification applications, where

- a specific set of one or more matching algorithms are known and available, or

- operating threshold is known and different from weights for Best Practice NFIQ summarization.

some examples :: quality monitoring

		#samples	#NFIQ 1	#NFIQ 2	#NFIQ 3	#NFIQ 4	#NFIQ 5	NFIQ summary
across consulates	1	1431	939	247	84	12	3	**98.13**
	2	4768	831	197	184	39	15	**95.50**
	3	1770	757	234	223	70	35	**92.82**
across airports	J	795	197	322	240	25	11	**92.80**
	L	929	283	358	255	18	15	**93.42**
	M	1346	473	536	302	22	13	**94.74**
per user	1	20	4	6	0	1	9	**52.24**
	2	20	16	4	0	0	0	**99.48**

more examples :: metadata analysis

age	younger than 30	between 30 and 65	older than 65
NFIQ summary	98.20	95.60	85.40

Data: poe-bva :: 183449 right index

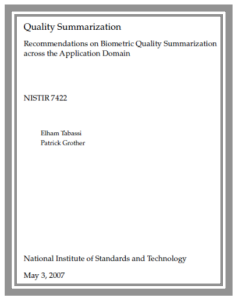
gender	ALL	male	female
NFIQ summary	94.93	95.98	93.80

Data: poe-bva :: 183449 left index

summary

- enterprise-wide quality summarization is operationally useful
- summarized quality value should be predictive of performance
  - for verification, should serve as measures of the overall expected false non-match rate
- NIST has established procedures for NFIQ summarization

Thank You


Quality Summarization

Recommendations on Biometric Quality Summarization across the Application Domain

NISTIR 7422

Elham Tabassi
Patrick Grother

National Institute of Standards and Technology

May 3, 2007

The quality summarization report is on line
www.itl.nist.gov/iad/894.03/quality/reports/enterprise.pdf

NIST Quality Program
www.itl.nist.gov/iad/894.03/quality/index.html

Feedback is welcomed.